The *Streptomyces lividans* Family 12 Endoglucanase: Construction of the Catalytic Core, Expression, and X-ray Structure at 1.75 Å Resolution[†,‡]

Gerlind Sulzenbacher,[§,||] François Shareck,[⊥] Rolf Morosoli,[⊥] Claude Dupont,[⊥] and Gideon J. Davies*[,§]

Department of Chemistry, University of York, Heslington, York YO1 5DD, England, European Molecular Biology Laboratory (EMBL), DESY, Notkestrasse 85, 22603 Hamburg, Germany, and Centre de Recherche en Microbiologie Appliquée, Institut Armand-Frappier, Université du Québec, Laval-des-Rapides, C. P. 100, Québec, Canada

Received September 29, 1997; Revised Manuscript Received November 7, 1997[⊗]

ABSTRACT: Cellulases are the glycoside hydrolases responsible for the enzymatic breakdown of the structural plant polymer cellulose. Together with xylanases they counteract the limitless accumulation of plant biomass in nature and are of considerable fundamental and biotechnological interest. Endoglucanase CelB from *Streptomyces lividans* performs hydrolysis of the β-1,4-glycosidic bonds of cellulose, with net retention of anomeric configuration. The enzyme is a member of glycoside hydrolase family 12 [Henrissat, B., and Bairoch, A. (1996) *Biochem. J. 316*, 695−696], which had previously eluded detailed structural analysis. A truncated, but catalytically competent form of CelB, lacking the flexible linker region and cellulose-binding domain, has been constructed and overexpressed in a *S. lividans* expression system. The three-dimensional X-ray structure of the resulting catalytic domain, CelB2, has been solved by conventional multiple isomorphous replacement methods and refined to an *R* factor of 0.187 at 1.75 Å resolution. The overall fold of the enzyme shows a remarkable similarity to that of family 11 xylanases, as previously predicted by hydrophobic clustering analysis [Törrönen, A., Kubicek, C. P., and Henrissat, B. (1993) *FEBS Lett. 321*, 135−139]. The 23 kDa protein presents a jelly-roll topology, built up mainly by antiparallel β-sheets arranged in a sandwich-like manner. A deep substrate-binding cleft runs across the surface, as has been observed in other endoglucanase structures, and is potentially able to accommodate up to five binding subsites. The likely catalytic nucleophile and Brønsted acid/base, residues Glu 120 and Glu 203, respectively, have their carboxylate groups separated by a distance of approximately 7.0 Å and are located approximately 15 Å from one end of the cleft, implying a −3 to +2 active site.

Cellulases are produced by cellulolytic bacteria and fungi and secreted into the extracellular environment, where they depolymerize high molecular weight cellulose into smaller oligo- and disaccharides, which can be used as a carbon and energy source. Cellulose is a polymer of β-1,4-linked D-glucosyl units, rendered highly crystalline by strong associative forces established between the resulting single polymer chains. It is therefore extremely resistant to enzymatic attack (*1*). Cellulolytic micro-organisms overcome the difficult task of cellulose breakdown by producing batteries of different catalytic and noncatalytic enzyme domains acting in combination on the substrate and displaying multiple types of synergy (for review see ref 2). These enzymes usually consist of a discrete catalytic core, which is connected to a cellulose binding domain (CBD)[1] , via a flexible linker. In some anaerobic organisms, the structural organization is more elaborate, and several catalytic, cellulose-binding and docking domains are organized as a multifunctional supermolecular complex known as the cellulosome (*3*). The catalytic domains of cellulases and xylanases have been classified into 13 distinct glycosyl hydrolase families on the basis of sequence similarities (*4*−*7*). The families may display different protein architectures, substrate specificity, *exo/endo* preference, and reaction stereochemistry, but all hydrolyze the glycosidic bond utilizing a general acid/base catalysis, essentially as outlined by Koshland (*8*), which may lead to either an inversion or a net retention of the anomeric configuration.

The *Streptomyces lividans* endoglucanase CelB is a member of glycoside hydrolase family 12 for which no structure has yet been determined, despite preliminary reports on fungal enzymes from *Aspergillus aculeatus* (*9*, *10*) and *Trichoderma reesei* (*11*). Family 12 enzymes are known to catalyse glycosidic bond cleavage with net retention of anomeric configuration (*12*−*14*), via a double-displacement mechanism in which a glycosyl-enzyme intermediate is formed and subsequently hydrolysed via oxocarbenium-ion transition states (*15*). Since protein structure is often more highly conserved than sequence, some of the families show distant but significant structural relationships hinting at a common ancestry. In the case of the family 12 cellulases,

[†] This work was funded in part, by the Biotechnology and Biological Sciences Research Council (BBSRC), the European Union (contract BIO4-CT97-5053), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Fonds pour la formation de chercheurs et l'aide à la recherche du Québec (FCAR). G.J.D. is a Royal Society University Research Fellow.

[‡] Coordinates for the structure described in this paper have been deposited with the Brookhaven Protein Data Bank (accession reference: 1NLR.PDB).

* Corresponding author and reprint request. Phone: 44-1904-432596. Fax: 44-1904-410519. E-mail: davies@yorvic.york.ac.uk.

[§] University of York.

[||] EMBL Hamburg.

[⊥] Institut Armand-Frappier.

[⊗] Abstract published in *Advance ACS Abstracts,* December 15, 1997.

[1] Abbreviations: CelB (CelB2), family 12 endoglucanase from *Streptomyces lividans*; CBD, cellulose binding domain; EGI, family 7 endoglucanase from *Fusarium oxysporum*; HCA, hydrophobic cluster analysis; GH-C glycoside hydrolase clan C; XYNII, family 11 xylanase II from *Trichoderma reesei*.

S0006-2960(97)02407-0 CCC: $14.00

hydrophobic cluster analysis indicated significant structural similarity with the xylanases from family 11 (*16*). In this paper, we describe, in detail, the first three-dimensional structure of an endoglucanase from family 12. The folding motif and catalytic machinery do indeed display an astonishing similarity to the family 11 xylanases. Preliminary analysis of the active site structure indicates the likely molecular basis for the difference in substrate specificity between these two families.

MATERIALS AND METHODS

PCR Templates, Primers, Conditions, and Cloning. The gene encoding the truncated form of CelB was constructed by PCR (*17*) using plasmids pIAF9 and pIAF18 as templates (*18*, *19*) and the following primers: CB-XC, 5′-TTCGCAGATCGTGGTGTCGGCGTGTGCCCCCGGCGG-3′; CB-XN, 5′-CCGCCGGGGGCACACGCCGACACCACGATCTGCGAA-3′; xlnA5-*Sph*I, 5′-GGGGCATGCGCTCCTACGCCCTTCCCAGATCAGG-3′; and celB234-*Sac*I, 5′-CGGGAGCTCTCAGCCCGGGTCGCCGGGGTCGGTGCC-3′.

All amplifications were performed using a GeneATAQ controller (LKB-Pharmacia) in 100 μL reactions with 1 ng of template, 50 pmol of each primer, 4% propionamide, 0.05% Tween20, 20 nmol of each deoxynucleotide and 2.5 units of *Pfu* polymerase (Stratagene) and the reaction buffer supplied by the manufacturer. The first PCR cycle consisted of a denaturation step at 95 °C for 5 min, an annealing step at 55 °C for 5 min, and a polymerization step at 72 °C for 3 min. Then, DNA was amplified by 30 cycles: 94 °C for 1 min, 55 °C for 1 min, and 72 °C for 3 min, followed by an extension of the unfinished products at 72 °C for 7 min. A sample (10 μL) was analyzed by agarose gel electrophoresis and the amplicons were recovered using the High Pure PCR Product Purification Kit (Boehringer Mannheim). Protoplasts of *S. lividans* 10−164 (*msi*K$^-$) (*20*) were transformed and regenerated on R5 medium according to ref 21. Transformants were grown for 24 h on Stewart solid medium containing 1% carboxymethyl cellulose (CMC) (*22*) and stained with Congo red (*23*). All constructions were verified by sequencing using the ALF automatic sequencer (Pharmacia Biotech).

Enzyme Production and Purification. Seven-day-old cultures of *S. lividans* from Bennett-thiostrepton plates were used as initial inoculum. The spores were scraped from the plates and inoculated into 12.5 mL minimal M15 medium (composition per liter: glucose, 10 g; K_2HPO_4, 5.0 g; $NaNO_3$, 1.4 g; KH_2PO_4, 1.0 g; $CaCl_2$, 300 mg; $MgCl_2$, 300 mg; $FeSO_4 \cdot H_2O$, 5.0 mg; $CoCl_2 \cdot 6H_2O$, 2.0 mg; $MnSO_4 \cdot H_2O$, 1.6 mg; $ZnSO_4 \cdot 7H_2O$, 1.4 mg; Tween 80, 2.0 mL; pH 7.3) and incubated for 18 h at 34 °C with agitation. Bacteria were recovered by centrifugation, used to inoculate 500 mL of the same medium and allowed to grow for 72 h under the same conditions. Proteins contained in the supernatant were first concentrated by ultrafiltration with a 3 kDa cut-off membrane (Omega). The concentrated proteins were then precipitated with 70% ammonium sulfate and, after centrifugation, the precipitate was dissolved in 50 mM sodium citrate buffer, pH 6.5. Samples of 200 mg of protein were then loaded on a phenyl Sepharose column (2.5 × 30 cm; Pharmacia) in 20 mM piperazine buffer containing 200 mM NaCl and 1 M ammonium sulfate, pH 6.0. Proteins were eluted with a decreasing linear gradient to 0 M ammonium sulfate, followed by an increasing linear gradient to 50% ethylene glycol. The CelB2 containing fractions were pooled, dialyzed against Milli-Q water and freeze-dried. The second step consisted in separation on a Superdex HR75 beaded column (3 × 60 cm; Pharmacia) with 100 mM sodium phosphate pH 6.0, as the eluant. Purified CelB2 containing fractions were pooled, dialyzed, and freeze-dried. Protein concentration was determined according to (*24*) using bovine serum albumin as standard (Bio-Rad). Determination of protein purity was assessed by SDS−PAGE (*25*) and Western blot analysis with anti-CelB antibodies (*26*). Analytical isoelectric focusing was carried out on PhastGel (Pharmacia) containing Pharmalyte carrier ampholytes from pH 3 to 9 using the automated PhastSystem (Pharmacia). The gels were silver-stained after the run (Separation and Development Technique Files no. 100 and 210; Pharmacia-LKB Biotechnologies, Uppsala, Sweden).

Crystallization and Data Collection. Crystals were obtained by the hanging-drop vapor diffusion method from 15 mg mL^{-1} enzyme solution in acetate buffer at pH 4.5 and using 30% (w/w) PEG 1500 as precipitant. Wedge-shaped crystals grew to a maximum size of 0.3 × 0.2 × 0.1 mm within 1−2 days. Crystals belong to space group $P2_12_12$, with cell dimensions $a = 48.49$ Å, $b = 95.48$ Å, and $c = 40.52$ Å. There is one molecule in the asymmetric unit, suggesting a solvent content of approximately 40%. Diffraction data from a single native crystal were measured to 1.70 Å resolution (referred to as high resolution native data set) on a Raxis II image plate system with a Cu rotating anode source. Cryocrystallographic techniques were employed using a solution of 50% (w/w) PEG 1500, buffered at pH 4.5, as cryoprotectant. An extended search for heavy-atom derivatives was carried out using standard soaking methods. All derivative data sets were collected at 120 K on a MARresearch Image Plate scanner mounted on a rotating anode generator with CuKa radiation. A second native data set, to 2.25 Å resolution (hereafter low-resolution data set) was collected on a crystal originating from the same crystallization batch and on the same imaging plate system as for the derivatized crystals, in order to reduce systematic errors. All data were processed and reduced with DENZO and SCALEPACK (*27*), and all subsequent computing used the CCP4 (*28*) suite of programs unless otherwise stated.

Structure Determination and Refinement. The positions of the heavy atoms were determined by inspection of the isomorphous difference Patterson maps and confirmed by analysis of the anomalous difference Patterson maps. The heavy atom parameters were refined with the program MLPHARE, utilizing both isomorphous and anomalous differences. Cross-phase difference Fourier maps were calculated to locate further heavy-metal sites in a total of six derivatives (mercury acetate, methyl mercury chloride, uranyl acetate, platinum terpyridyl chloride, samarium acetate, and potassium tetrachloroplatinate). Automated solvent flattening, assuming a solvent content of 33%, was carried out with the program DM (*29*). All model building was carried out with the program O (*30*). Least-squares refinement was carried out with the Maximum Likelihood program REFMAC, incorporating bulk solvent corrections and anisotropic F_{obs} *vs* F_{calc} scaling (*31*) and utilizing experimental phases. Five percent of the reflections were set aside during refinement and map calculation for cross-validation purposes, and the free R was used to monitor

refinement strategies (*32*). The solvent structure was built by automatic insertion and deletion of water molecules by the program ARP (*33*). The stereochemistry of the model was verified with the program PROCHECK (*34*) and secondary structure assigned using DSSP (*35*). The atomic coordinates of the structure have been deposited with the Brookhaven Protein Data Bank (*36*), with accession reference 1NLR.

RESULTS AND DISCUSSION

Construction, Purification, and Characterization of the Truncated Form of CelB. Since glycanases bearing a substrate binding domain linked by a flexible region rarely prove suitable for crystallization, the CBD of the native CelB was removed at the genetic level by PCR. In addition, the native signal peptide of the CelB gene contains two leucine residues encoded by rarely used TTA codons, which could hamper the efficient translation of the mRNA into protein (*18*). In order to optimize the translation and efficient secretion of CelB, its signal peptide sequence was replaced by the one from the *S. lividans* xylanase A gene, which was amplified using primers xlnA5-*Sph*I, CB-XN, and pIAF18 as template. The truncated CelB (hereafter referred to as CelB2) was constructed using primers CB-XC, celB234-*Sac*I, and pIAF9 as template. The two amplicons were then joined together using primers xlnA5-*Sph*I and celB234-*Sac*I, resulting in a chimeric protein without the carboxy-terminal CBD. This CBD was identified by sequence homology with other CBDs and belongs to family IIa (*37*). The primers xlnA5-*Sph*I and CelB234-*Sac*I allowed the incorporation of a restriction site at the 5′ end, a TGA stop codon, and a unique *Sac*I restriction site at the 3′ end after the amino acid G234. This residue is located at the far end of the linker region separating the N-terminal catalytic domain from the CBD. The incorporation of unique restriction sites at both ends of the PCR product allowed directional subcloning of the DNA fragment into the plasmid pIAF20−109 resulting in plasmid pIAF933 (Figure 1). Sequencing of the final construct revealed divergences with the original sequence at positions 153 and 154 where glycine residues were found instead of alanine. Verification of the original template (pIAF9) indicated that these divergences were not due to PCR amplification misincorperations but to ambiguities in the previous DNA sequencing. Thiostrepton-resistant transformants showing cellulase activity on Stewart-CMC medium were isolated. The secreted, truncated enzyme was purified as described and showed mobility on SDS−PAGE corresponding to the expected protein of approximately 22 kDa and reacted with anti-CelB antibodies in Western blot analysis (data not shown). The pI of the CelB2 enzyme was estimated to 3.9 compared to 4.2 for the full-length CelB enzyme. No significant differences were observed for the physicochemical or catalytic properties between both forms of the enzyme using CMC as substrate (data not shown).

3-D Structure Determination and Quality of the Final Model Structure. A summary of the data collection statistics for the native data is given in Table 1. Despite the high quality of diffraction data and reasonable isomorphous differences, it was not possible to locate any heavy atom positions in the difference Patterson maps calculated with derivative data scaled to the high-resolution native data. Using the second native, low-resolution data set, a strong single peak for the UO_2Ac_2 derivative could be unambiguously identified in all the Harker sections and confirmed in the anomalous difference Patterson maps, suggesting systematic errors between different detectors or nonisomorphism between the different batches of crystals. The final MIR data had a figure of merit of 0.778 for centric reflections and 0.712 for acentric reflections. Automated flattening and phase extension, to 1.75 Å resolution, gave an easily interpretable electron density map. The final $2F_{obs} - F_{calc}$

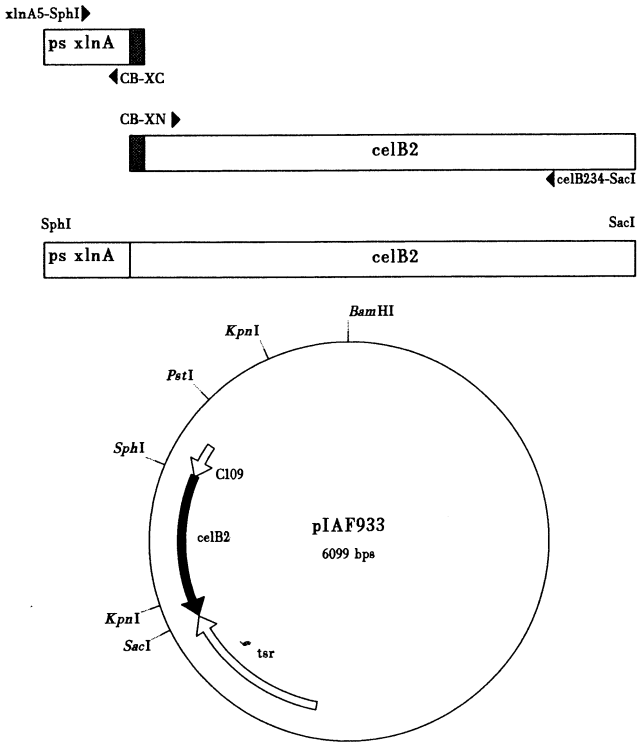

FIGURE 1: Construction of plasmid pIAF933. The signal peptide sequence of xlnA was fused to the catalytic domain of celB by PCR. The primers xlnA5-*Sph*I and CB-XC were used to amplify the signal peptide of xylanase A and the catalytic domain of the cellulase B was amplied using primers CB-XN and celB234-*Sac*I. The filled boxes represents respectively the primers CB-XC and CB-XN which overlap the 3′ end of the signal peptide and the 5′ end of cellulase B. The two amplicons were mixed together in a PCR reaction using the primers xlnA5-SphI and celB234-SacI. The final amplicon containing the fusion was digested with restriction enzymes *Sph*I and *Sac*I and cloned downstream of the promoter C109.

Table 1: Refinement and Structure Quality Statistics for the *Streptomyces lividans* CelB2 Structure

resolution of data (outer shell) (Å)	20−1.75 (1.84−1.75)
R_{merge}[a] (outer shell)	0.047 (0.287)
completeness (outer shell) (%)	93.7 (67.9)
multiplicity (outer shell)	6.77 (2.78)
no. protein atoms (residues 1−222)	1664
no. solvent waters (B < 50 Å^2)	200
resolution used in refinement	15.0−1.75 Å
R_{cryst}[b]	0.187
R_{free}	0.240
rmsd 1−2 bonds (Å)	0.012
rmsd 1−3 angles (Å)	0.030
rmsd chiral volumes (Å^3)	0.139
avg mean B (protein atoms (Å^2)	18.9
avg main chain B (Å^2)	18.8
avg side chain B (Å^2)	21.3
avg solvent B (Å^2)	33.2
main chain ΔB, bonded atoms (Å^2)	2.2

[a] $R_{merge} = \sum_{hkl}\sum_i|I_{hkli} - \langle I_{hkli}\rangle| / \sum_{hkl}\sum_i \langle I_{hkli}\rangle$. [b] $R_{cryst} = \sum||F_o| - |F_c||/\sum|F_o|$.

map, calculated with maximum likelihood weighting, shows no discontinuity in the electron density, except for the flexible loop Gly 153-Asn 158. This loop has main-chain temperature factors around 40−50 Å^2 and may exhibit dynamic disorder in the crystal. Maximum likelihood refinement proceeded smoothly and the high quality of the experimental phases justified their use to improve convergence. The final model contains 1664 nonhydrogen protein atoms and 200 water molecules, and the final crystallographic *R* factor and *R* free, for the resolution range 15.0−1.75 Å, are 0.187 and 0.240. The current model displays good stereochemistry with root-mean-square deviations from stereochemical target values of 0.012 Å and 1.7° for 1−2 bonding distances and 1−3 angles, respectively. Main chain atoms and side chain atoms display an average *B* factor of 18.4 Å^2 and 21.1 Å^2, respectively. A total of 90.1% of the nonglycine residues have their conformational angles (ϕ/ψ) in the most favored regions of the Ramachandran plot (*38*) with no residues in the disallowed regions, as defined by PROCHECK (*34*).

Description of the Overall Structure. The catalytic domain of CelB2 folds into a single domain with approximate dimensions of 40 × 40 × 30 Å and consists almost exclusively of β-strands and interconnecting loops. Two antiparallel β-sheets, A and B, consisting of six and nine strands, respectively, sit on top of each other, with their hydrophilic faces exposed to solvent and their hydrophobic faces packed against each other in a sandwich like manner (Figure 2). Both sheets are curved and give rise to a long open cleft on one side of the protein, as it is expected from an enzyme known to have endocatalytic activity. This molecular architecture is the classic jelly-roll topology found in the plant legume lectins, such as concanavalin A (*39, 40*) and which has been observed in a number of glycoside hydrolase structures such as the bacterial β-1,3-1,4-glucanase from family 16 (*41*), the family 7 cellulases (*42*−*45*), and the family 11 xylanases (for example refs 46 and 47). In particular, there is a striking resemblance of the CelB2 structure to that displayed by the family 11 xylanases. This confirms early predictions based on hydrophobic cluster analysis (*16*) and will be discussed further below.

Seven strands of β-sheet B, numbered B1−B7, form the center and the sides of the 35 Å long substrate binding cleft. β-Sheet A forms the back of the protein, together with the remaining two strands of sheet B, B8 and B9, an α-helix and an α-helical turn. The six strands of β-sheet A are shorter, and two of them, A3 and A4, are split into several shorter segments, conferring the curvature necessary to follow sheet B. Hydrophobic packing between the sheets involves Phe 125 and the invariant residues: Trp 148, Phe 163, and Phe 179. The polypeptide chain passes several times between the two sheets with the resulting loops forming the rim of the cleft and incorporating a number of conserved conformation-liberal glycines. A long loop is inserted between strands B3 and A5 and contains both a short 3_{10} helix and two additional, β-strands, C1 and C2. These two strands are very short and make up one of the rims of the active site cleft, together with the two loops between strands B5 and B6, and strands B7 and B8. An additional long surface loop is inserted between β-strands B6 and B9 and closes the substrate binding canyon at its far end. This structural feature is also conserved in the family 11 xylanases and has been referred to as the cord (*46*). It has been suggested that this loop may undergo conformational changes upon substrate binding, with conformation flexibility accompanied by a prolyl *cis-trans* isomerisation (*46*). In the family 12 sequences, the corresponding amino acids are highly conserved, with the invariant Pro 133 tightly packed between the aromatic side chains of Trp 124 and Trp 151, which could interact with the hydrophobic faces of the oligosaccharide ligand should a loop movement occur.

The 35 Å long active site cleft of CelB2 is about 15 Å deep and 8 Å wide for most of its extension. By analogy with the *Trichoderma reesei* xylanase II, for which the directionality of substrate binding is well-known (*48*), and using the molecular model of cellotetraose (*49*) as a ruler, we have been able to establish the possible number of substrate binding sites for CelB2. The enzyme is likely to bind five, possibly six, glucopyranose units in its active site cleft, occupying subsites −3 to +2, with cleavage taking place, by definition, between subsites −1 and +1 (nomenclature reviewed in ref 50).The substrate-binding cleft is lined with a number of aromatic residues which may bind the hydrophobic faces of the pyranoside rings as has been observed on many carbohydrate binding proteins (*51*). At the nonreducing end of the cleft, the side chains of Tyr 66, Phe 8, and His 65 are conserved or substituted by equivalent groups in other members of family 12 and are likely to interact with glucopyranose rings in either the −3 or the −2 subsite. Residues Asn 22 and Trp 24 are invariant and flank subsite −2. For a substrate in subsite −1, stabilization would occur through hydrogen bonding to the acid/base Glu 203, Asn 155, and Trp 24. All these groups are present or replaced by equivalent groups in the sequences of family 12 enzymes. The invariant Met 122 may function in hydrophobic stacking with the +1 subsite sugar. In the CelB2 structure, Met 122 has undergone chemical modification during either protein preparation or exposure of the crystal to X-rays and is present as a methionine sulphoxide. A methionine, Met 502, has been observed in an equivalent position in the active site of the unrelated family 2 β-galactosidases (*52*), but despite extensive studies on the enzyme system of *E. coli* β-galactosidase (*53, 54*), no role has yet been assigned to this residue. For the special case of subsite +2, further crystallographic studies of enzyme−oligosaccharide complexes will be required, in order to determine whether the corresponding glucopyranose unit binds loosely to the rim of the groove, or if a conformational change of the cord permits tight binding deep down at the bottom of the groove, as suggested by Törrönen and co-workers (*46*).

The structure contains a single *cis*-proline, Pro 76, forming a turn in the extended loop region between β-strands B5 and A3. This residue is not conserved, although an equivalent *cis*-proline, Pro 51, is present in the *Trichoderma reesei* xylanase II structure. Two disulphide bridges, located between residues Cys 5 and Cys 31 and Cys 64 and Cys 69, stabilize the 3D-fold by connecting strand A1 to A2 and C1 to C2. Although the molecule contains 19 acidic and 11 basic residues, only three salt bridges could be observed. These are established between Arg 78 and Asp 80, Arg 112 and Asp 108, and between Asp 178 and Arg 181, these last two residues being located on the α-helix. The single lysine, Lys 55, in the sequence plays a crucial role and may stabilize the structure by making strong hydrogen bonds with the main chain carbonyl oxygens of Ala 44, Asp 45, and Asn 207.

Comparison with Family 11 Xylanases. It was predicted that the family 12 cellulases were related to family 11

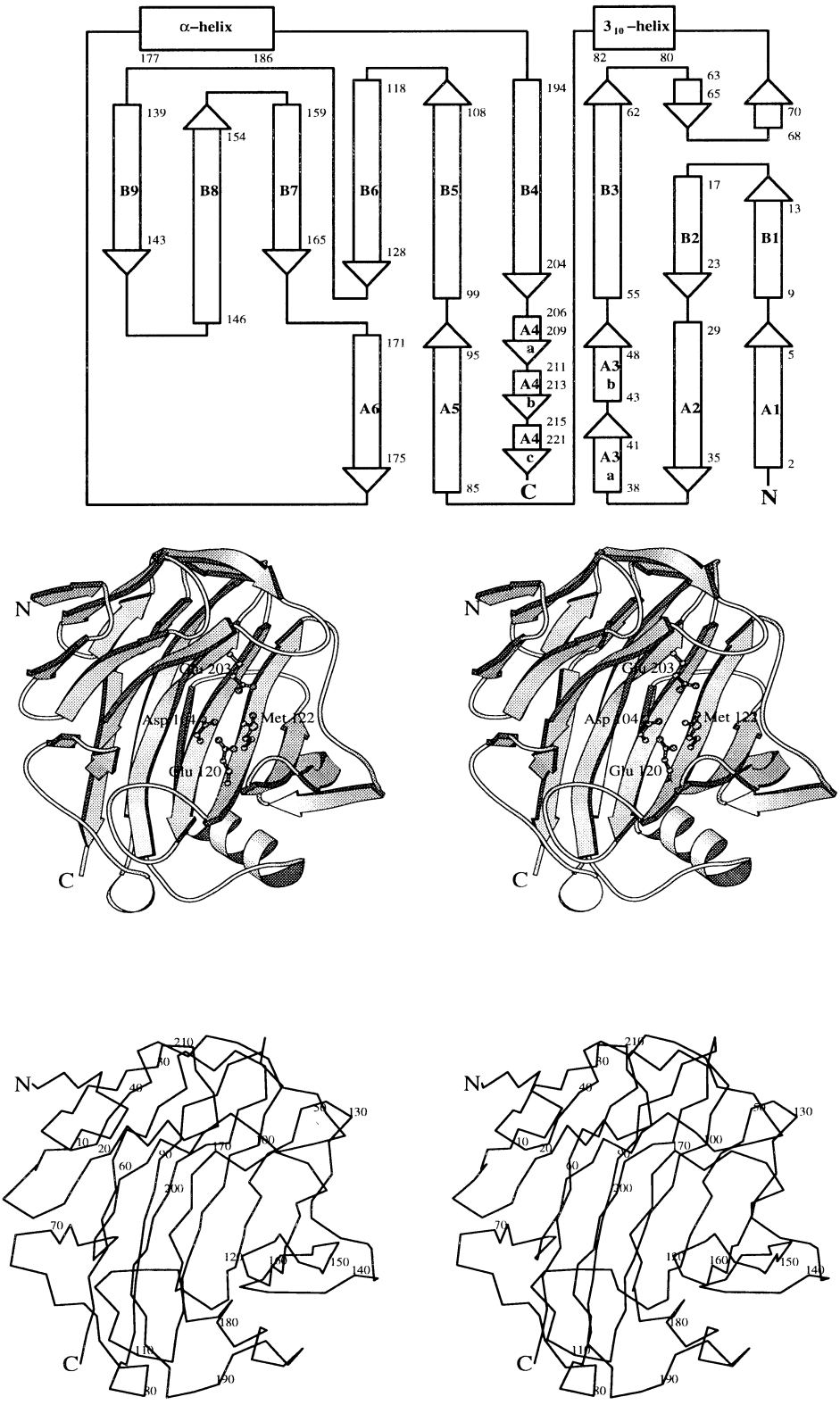

FIGURE 2: Structure of the catalytic core of the family 12 endoglucanase from *S. lividans*. Top, topology diagram, helices are shown as rectangles and sheets as arrows. Middle, divergent stereo cartoon representation, residues mentioned in the text are shown with ball-and-stick representation. Bottom, stereo α-carbon trace with every 10th residue indicated. The latter two representations, and those in Figures 3 and 4, were prepared with the MOLSCRIPT program (*63*).

xylanases on the basis of both conventional sequence comparisons (*55*) and hydrophobic cluster analysis (*16*). Although the sequence similarity between the two families is extremely low, HCA indicated that the structures would have the same overall fold but with significant structural divergence. Henrissat has coined the term clan to describe groupings of families that display significant similarities in tertiary structure together with conservation of the catalytic residues and mechanism and which may therefore be evolutionarily related (*6, 7*). On the basis of the above results, families 11 and 12 were proposed to form a clan, GH-C, despite the fact that there was no known family 12 structure. These proposals are unambiguously confirmed by the structure presented here. An overlap of the CelB2

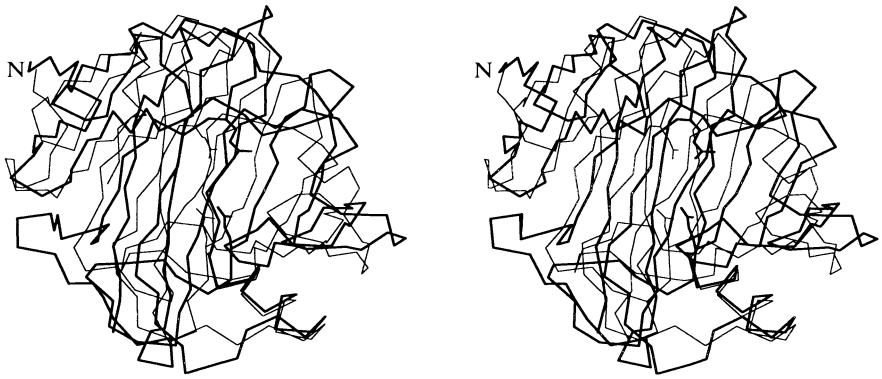

FIGURE 3: Stereo overlap of the *S. lividans* family 12 endoglucanase (bold) with the family 11 xylanase from *T. reesei* (faint).

structure with the Xylanase II, XYNII, from *Trichoderma reesei* (PDB-entry 1RED) reveals that the protein core, consisting in both cases of a 6- and a 9-stranded β-sheet, is extremely similar in the two enzymes, whereas significant differences are observed in the surface-loop regions (Figure 3). This is reflected in an rmsd for 153 equivalent Ca positions of 1.9 Å, using a simple rigid-body least-squares fitting algorithm. CelB2 contains 32 residues more than XYNII, and its active site cleft is about 7 Å longer. A common feature to all published family 11 xylanase structures is the thumb, a long loop-extension overlooking the active site. This structural feature is less pronounced in CelB2, where the place of the thumb is occupied by the two surface loops between strand B5-B6 and B7-B8, the later being the flexible loop, Gly 153-Asn 158. The extension between strand B3 and A5 in CelB2, which contains the small β-sheet C and the 3_{10} helix, is missing in XYNII. In CelB2, β-sheet C forms one side of the groove at the reducing end, corresponding to the putative substrate binding sites -4 and -3, which are absent in the XYNII structure. At the reducing end of the groove, near the putative subsite $+2$, the surface loop between β-strand A3b and B3 is longer than the corresponding loop in XYNII, which has the effect that in this point the groove is much deeper in CelB2 with respect to XYNII. The catalytic residues, Glu 120 and Glu 203 in CelB2 (equivalent to Glu 86 and Glu 177 in XYNII), are invariant in all family 11 and family 12 enzymes and are located on topologically equivalent β-strands: strand B6 for the nucleophile and strand B4 for the acid/base.

The Active Site and Catalytic Mechanism. Family 12 enzymes perform hydrolysis with a net retention of anomeric configuration (*12*−*14*). This is widely believed to be by a double-displacement reaction mechanism in which a covalent glycosyl-enzyme intermediate is formed and subsequently hydrolysed via oxocarbenium-ion transition states, essentially as described by Koshland (*8*, and, for review, refs 15 and 56). The striking similarity between the family 11 xylanases and this family 12 member allows us to assign roles for the respective catalytic residues. In family 11, the enzymatic nucleophile has been trapped through the use of 2-deoxy-2-fluoro-β-xylobiosides with reactive leaving groups (*57*). The corresponding residue in the family 12 structure CelB2 is Glu 120. Geometric considerations and pH titrations with ^{13}C-labeled enzymes revealed that Glu 172 functions as the catalytic acid/base in the family 11 *Bacillus circulans* xylanase (*58*). This corresponds to Glu 203 in CelB2. The mechanisms whereby the appropriate pK_as are maintained for the catalytic nucleophile and the acid/base in the family 11 xylanases involves an elegant pK_a cycling, in which the microscopic pK_a of each group is dependent on the ionization state of the other. Hence, upon formation of the covalent glycosyl-enzyme intermediate, there is a change in the pK_a of the catalytic acid/base, allowing it to function as a Brønsted base for the deglycosylation half-reaction. The similarity between the family 11 xylanases and the family 12 cellulases suggests that an equivalent pK_a cycling is also likely to occur in family 12. In the CelB2 structure, the two catalytic glutamates, Glu 120 and Glu 203, sit in front of each other at the bottom of the substrate-binding groove, with their carboxylate groups separated by approximately 7 Å (Figure 4). Such a distance is longer than the 5.5 Å normally encountered in retaining enzymes (*56*, *59*), and it seems likely that conformational change would be required for concomitant protonation of the leaving group and nucleophilic attack at C1, a constriction of the enzyme structure upon ligand binding having previously been observed in the related xylanase structures (*48*). The distant separation of these residues may simply reflect their protonation states at pH 4.5 or may suggest that a substrate-induced conformational change is required in order to position the residues for catalysis as has been observed in a family 5 endoglucanase structure (*60*). The two catalytic residues are flanked at one site by a third acidic group, Asp 104, which is conserved or substituted by a glutamate in other family 12 enzymes, and on the other side by the invariant Met 122. The carboxyl function of Asp 104 is about 3 Å away from the nucleophile and 5 Å away from the acid/base, but in an orientation which makes hydrogen bonding to the catalytic glutamates unlikely. There is no equivalent to Asp 104 in the related family 11 xylanase sequences which have an invariant tyrosine in this position. A distant, but possibly significant, relationship can also be seen with the family 7 cellulases such as the *Fusarium oxysporum* EGI (*42*, *43*). Both EG I and CelB2 are formed by two large anti-parallel β-sheets. In CelB2, the β-sheets are both concave, but in EG I, the two sheets stack on with their convex faces on top of each other, and expose the hydrophilic concave faces. The structure of EGI is almost twice as big as CelB2, 400 residues against 222, and has a cleft of approximately 50 Å, against the length of approximately 35 Å observed for CelB2. In EGI, the catalytic residues are Glu 197 and Glu 202, which play the role of nucleophile and general acid/base, respectively. A third acidic group, Asp 199, is in intimate contact with the two glutamates and may be important for maintenance of the pK_a of the proton donor, as has been demonstrated on other three-acid systems

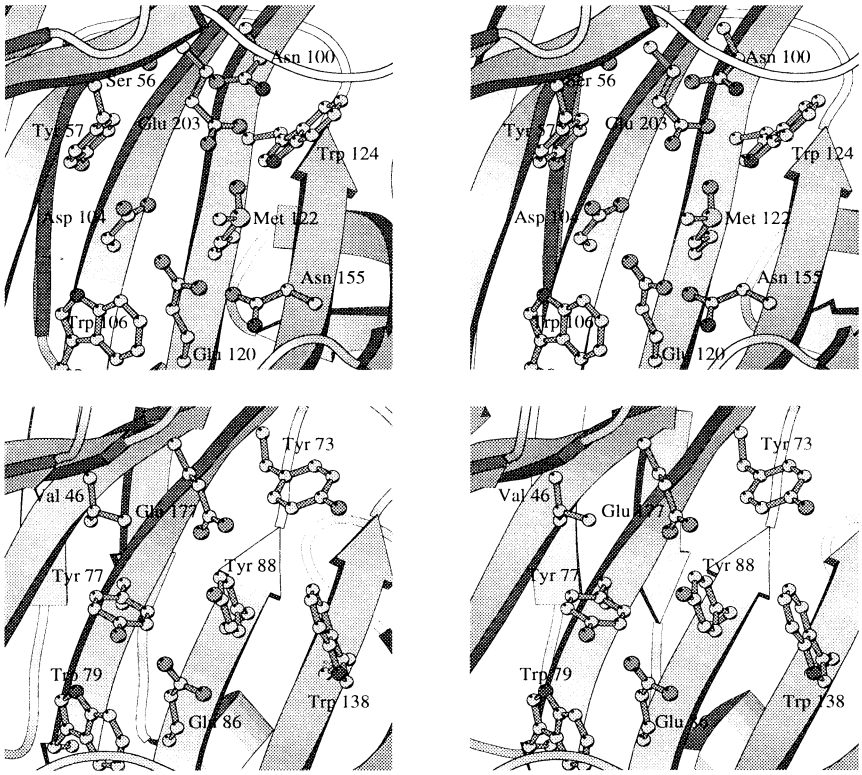

FIGURE 4: Active site environments for CelB2 (top) and the family 11 xylanase from *T. reesei* (bottom).

(for example, refs 61 and 62). This constellation of three acidic groups is superficially similar to the three groups in CelB2, but in CelB2, these residues are all donated from different structural elements, and the third aspartate lies in a structurally nonequivalent location. The reason for the presence of a third acid group in this cellulolytic enzyme system, but not in the related family 11 xylanases, is intriguing. The simplest explanation is that this residue functions in binding the exocyclic hydroxyl of the glucose rings, but more subtle explanations involving the slightly different chemistry of xylose compared to glucose are also possible.

At the catalytic level, the family 11 xylanases are very specific towards xylan whereas the preferred substrate for family 12 cellulases is cellulose, although notable activity can also be observed on xylan (*55*). The main difference between the two natural substrates can be found in the C5 position of the glucopyranose ring, where xylan is missing the exocyclic hydroxy-methyl substituent. In CelB2 the side chain of Tyr 57, neighboring the proton donor Glu 203, points away from the putative −1 subsite and is locked in this position by a hydrogen bond between its OH and the main chain carbonyl of Ser 102. The result is the formation of a cavity adjacent to the Glu 203, which could accommodate the exocyclic hydroxy-methyl substituent of glucose. Hydrogen-bonding functionalities available to a potential O-6 group In XYNII, by contrast, the equivalent position is occupied by an invariant valine, Val 46, whose hydrophobic side chain points towards the proton donor Glu 177. Not only would this probably favor binding of the more hydrophobic xylose-based polymer, but modeling suggests that the exocyclic substituent of glucose would be prevented from binding in this position due to steric clashes.

Summary. In this work we present the first detailed description for a family 12 glycoside hydrolase, the endoglucanase CelB2 from *S. lividans*. The overall fold of the enzyme is extremely similar to that displayed by the family 11 xylanases, confirming the predictions based on hydrophobic cluster analysis. This similarity exists even at the level of the catalytic machinery, which is completely invariant between the two families of enzyme, but subtle differences in the hydrophobic nature of the −1 subsite suggest the likely determinant of substrate specificity. A distant similarity to the family 7 cellulases manifests itself in an approximate topological conservation as well as the existence of similarly located acidic residues. Further studies to establish the exact nature of substrate specificity in the family 11 and family 12 enzymes are currently in progress.

ACKNOWLEDGMENT

The authors would like to thank K. S. Wilson, E. J. Dodson, G. Murshodov, and D. Kluepfel for support and helpful discussion. We also like to acknowledge L. Trempe and L. Duval for their technical assistance and F. Bizouarn for providing some DNA constructs.

REFERENCES

1. Béguin, P., and Aubert, J.-P. (1993) *FEMS Microbiol. Rev. 13*, 25−58.
2. Gilbert, H. J., and Hazlewood, G. P. (1993) *J. Gen. Microbiol. 139*, 187−194.
3. Bayer, E. A., Morag, E., and Lamet, R. (1994) *Trends Biotechnol. 12*, 379−386.
4. Henrissat, B. (1991) *Biochem. J. 280*, 309−316.
5. Henrissat, B., and Bairoch, A. (1993) *Biochem. J. 293*, 781−788.
6. Henrissat, B., and Bairoch, A. (1996) *Biochem. J. 316*, 695−696.
7. Henrissat, B., and Davies, G. J. (1997) *Curr. Opin. Struct. Biol. 7*, 637−644.
8. Koshland, D. E. (1953) *Biol. Rev. 28*, 416−436.

9. Okada, H. (1990) *Microbial Utilisation of renewable resources*, pp 1−7, Osaka University, Okada, Japan.
10. Hata, Y., Natori, K., Katsube, Y., Ooi, T., Arai, M., and Okada, H. (1994) *J. Mol. Biol. 241*, 278−280.
11. Ward, M., Wu, S., Dauberman, J., Weiss, G., Larenas, E., Bower, B., Rey, M., Clarkson, K., and Bott, R. (1993) in *Proceedings of the second TRICEL symposium on Trichoderma reesei cellulases and other hydrolases* (Suominen, P., and Reinikainen, T., Eds.) pp 153−158, Foundation for Biotechnical and Industrial Fermentation Research, Espoo, Finland.
12. Schou, C., Rasmussen, G., Kaltoft, M.-B., Henrissat, B., and Schülein, M. (1993) *Eur. J. Biochem. 217*, 947−953.
13. Schülein, M., Tikhomirov, D. F., and Schou, C. (1993) in *Proceedings of the second TRICEL symposium on Trichoderma reesei cellulases and other hydrolases* (Suominen, P., and Reinikainen, T., Eds.) pp 109−116, Foundation for Biotechnical and Industrial Fermentation Research, Espoo, Finland.
14. Schülein, M. (1997) *J. Biotechnol. 57*, 71−81.
15. Davies, G., Sinnott, M. L., and Withers, S. G. (1997) in *Comprehensive Biological Catalysis* (Sinnott, M. L., Ed.) pp 119−209, Academic Press, London.
16. Törrönen, A., Kubicek, C. P., and Henrissat, B. (1993) *FEBS Lett. 321*, 135−139.
17. Yon, J., and Fried, M. (1989) *Nucleic Acids Res. 12*, 4895.
18. Wittmann, S., Shareck, F., Kluepfel, D., and Morosoli, R. (1994) *Appl. Environ. Microbiol. 60*, 1701−1703.
19. Mondou, F., Shareck, F., Morosoli, R., and Kluepfel, D. (1986) *Gene 49*, 323−329.
20. Hurtubise, Y., Shareck, F., Kluepfel, D., and Morosoli, R. (1995) *Mol. Microbiol. 17*, 367−377.
21. Hopwood, D. A., Bibb, M. J., Chater, K. F., Kieser, T., Bruton, C. J., Kieser, H. M., Lydiate, D. J., Smith, C. P., Ward, J. M., and Schrempf, H. (1985) *Genetic manipulation of Streptomyces: a laboratory manual*, The John Innes Foundation, Norwich.
22. Shareck, F., Mondou, F., Morosoli, R., and Kluepfel, D. (1987) *Biotechnol Lett. 9*, 160−174.
23. Teather, R. M., and Wood, P. J. (1982) *Appl. Environ. Microbiol. 43*, 777−780.
24. Lowry, O. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J. (1951) *J. Biol. Chem 193*, 265−275.
25. Laemmli, U. K. (1970) *Nature 227*, 680−685.
26. Towbin, H., Staehelin, T., and Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4350−4354.
27. Otwinowski, Z., and Minor, W. (1997) in *Methods in Enzymology: Macromolecular Crystallography*, part A (Carter, C. W., Jr., and Sweet, R. M., Eds.) pp 307−326, Academic Press, London, New York.
28. Collaborative Computational Project Number 4. (1994) *Acta Crystallogr. D 50*, 760−763.
29. Cowtan, K. D., and Main, P. (1996) *Acta Crystallogr. D 49*, 148−157.
30. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A 47*, 110−119.
31. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D 53*, 240−255.
32. Brünger, A. T. (1992) *Nature 355*, 472−475.
33. Lamzin, V. S., and Wilson, K. S. (1993) *Acta Crystallogr. D 49*, 129−147.
34. Laskowski, R. A., McArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 282−291.
35. Kabsch, W., and Sander, C. (1983) *Biopolymers 22*, 2577−2637.
36. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. T., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.
37. Tomme, P., Warren, R. A. J., R C Miller, J., Kilburn, D. G., and Gilkes, N. R. (1995) in *Enzymatic degradation of insoluble carbohydrates: ACS Symposium Series* (Saddler, J. N., and Penner, M. H., Eds.) pp 142−163 (American Chemical Society, Washington, DC.
38. Ramakrishnan, C., and Ramachandran, G. N. (1965) *Biophys. J. 5*, 909−933.
39. Edelman, G. M., Cunningham, B. A., Reeke, G. N., Jr., Becker, J. W., Waxdal, M. J., and Wang, J. L. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 2580−2584.
40. Hardman, K. D., and Ainsworth, C. F. (1972) *Biochemistry 11*, 4910−4919.
41. Keitel, T., Simon, O., Borriss, R., and Heinemann, U. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 5287−5291.
42. Sulzenbacher, G., Driguez, H., Henrissat, B., Schülein, M., and Davies, G. J. (1996) *Biochemistry 35*, 15280−15287.
43. Sulzenbacher, G., Schülein, M., and Davies, G. J. (1997) *Biochemistry 36*, 5902−5911.
44. Kleywegt, G. J., Zou, J.-Y., Divne, C., Davies, G. J., Sinning, I., Ståhlberg, J., Srisodsuk, M., Teeri, T. T., and Jones, T. A. (1997) *J. Mol. Biol. 272*, 383−397.
45. Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K. C., Teeri, T. T., and Jones, A. (1994) *Science 265*, 524−528.
46. Törrönen, A., Harkki, A., and Rouvinen, J. (1994) *EMBO J. 13*, 2493−2501.
47. Wakarchuk, W. W., Campbell, R. L., Sung, W. L., Davoodi, J., and Yaguchi, M. (1994) *Protein Sci. 3*, 467−475.
48. Havukainen, R., Törrönen, A., Laitinen, T., and Rouvinen, J. (1996) *Biochemistry 35*, 9617−9624.
49. Raymond, S., Heyraud, A., Qui, D. T., Kvick, A., and Chanzy, H. (1995) *Macromolecules 28*, 2096−2100.
50. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) *Biochem. J. 321*, 557−559.
51. Vyas, N. K. (1991) *Curr. Opin. Struct. Biol. 1*, 732−740.
52. Jacobson, R. H., Zhang, X.-J., DuBose, R. F., and Matthews, B. W. (1994) *Nature 369*, 761−766.
53. Cupples, C. G., Miller, J. H., and Huber, R. E. (1990) *J. Biol. Chem. 265*, 5512−5518.
54. Richard, J. P., Huber, R. E., Lin, S., Heo, C., and Amyes, T. L. (1996) *Biochemistry 35*, 12377−12386.
55. Saarilahti, H. T., Henrissat, B., and Pavla, E. T. (1990) *Gene 90*, 9−14.
56. McCarter, J. D., and Withers, S. G. (1994) *Curr. Opin. Struct. Biol. 4*, 885−892.
57. Miao, S., Ziser, L., Aebersold, R., and Withers, S. G. (1994) *Biochemistry 33*, 7027−7032.
58. McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Körner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W., and Withers, S. G. (1996) *Biochemistry 35*, 9958−9966.
59. White, A., Withers, S. G., Gilkes, N. R., and Rose, D. R. (1994) *Biochemistry 33*, 12546−12552.
60. Dominguez, R., Souchon, H., Lascombe, M.-B., and Alzari, P. M. (1996) *J. Mol. Biol. 257*, 1042−1051.
61. Damude, H. G., Withers, S. G., Kilburn, D. G., Miller, R. C., Jr., and Warren, R. A. J. (1995) *Biochemistry 34*, 2220−2224.
62. Ståhlberg, J., Divne, C., Koivula, A., Piens, K., Claeyssens, M., Teeri, T. T., and Jones, T. A. (1996) *J. Mol. Biol. 264*, 337−349.
63. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

BI972407V